

Valutazioni su larga scala e ricerche in didattica della matematica:

esempi di studi condotti in Italia

XXXIX Seminario Nazionale di Ricerca in Didattica della Matematica ‘Giovanni Prodi’ RIMINI

Rimini, 24-25-26 marzo 2023

Federica Ferretti (Università di Ferrara), Giorgio Bolondi (Libera Università di Bolzano),
Alessandro Gambini (Sapienza Università di Roma), Chiara Giberti (Università di Bergamo), Alice
Lemmo (Università dell’Aquila), Francesca Martignone (Università del Piemonte Orientale),
Camilla Spagnolo (Libera Università di Bolzano)

1. Introduzione

Il seminario presenta un insieme di ricerche in didattica della matematica che a vario titolo fanno riferimento alle valutazioni nazionali su larga scala. In parallelo allo sviluppo delle più importanti esperienze internazionali (OCSE-PISA, IEA-TIMSS, IEA-PIRLS), negli ultimi decenni quasi la totalità dei paesi ha introdotto rilevazioni nazionali su larga scala degli apprendimenti degli studenti di scuola primaria e scuola secondaria (Eurydice, 2016; Looney, 2011; Morris, 2011). Queste rilevazioni, costruite con l’obiettivo di valutare l’apprendimento della matematica a livello di sistema, stanno avendo sempre più implicazioni dal punto di vista educativo, didattico, storico-culturale e politico sia a livello locale sia a livello globale (Atkin, 1998; Breakspear, 2012; Kanen, Morgan, & Tsatsaroni, 2014; Doig, 2006; Tasaki, 2017). Le valutazioni su larga scala hanno un forte impatto sistemico e sono state anche contestate per le loro possibili implicazioni pedagogiche, educative e politiche (si veda, ad esempio, Carnoy, 2015; Cochran-Smith, 2001). Tutto questo ha fatto sì che le valutazioni standardizzate su larga scala siano diventate, in diverse direzioni, oggetto di ricerche in didattica della matematica (De Lange, 2007; Kloosterman et al., 2015; Meinck, Neuschmidt, & Taneva, 2017; Suurtamm et al., 2016). All’interno di questo dibattito, la discussione verte sulla prospettiva con cui queste valutazioni/rilevazioni vengono utilizzate, rientrando in una problematica più ampia che coinvolge la valutazione sommativa in generale. Qualsiasi tipo di prova può essere utilizzata come strumento o come fine didattico. Noi abbracciamo la prospettiva per cui le prove, i dati che esse permettono di raccogliere e la loro elaborazione possano essere uno strumento utilizzabile per scopi didattici, educativi e di ricerca. In particolare, il gruppo di ricercatori che presenterà il seminario sta da anni indagando su come analizzare e utilizzare materiali e dati relativi alle valutazioni standardizzate nazionali italiane in ambito di ricerca, per vedere se e come possono essere utilizzate, in un’ottica di miglioramento dei processi di insegnamento e apprendimento della matematica. Come verrà approfondito durante il seminario, negli studi condotti le valutazioni standardizzate sono sia “oggetto di ricerca” sia “strumento di ricerca” per studiare specifici problemi di ricerca; per la formazione dei docenti di matematica, sia pre-servizio che in-servizio, aprendo così ulteriori direzioni di ricerca in un’ottica di un miglioramento dei processi di insegnamento-apprendimento fondato teoricamente e sperimentalmente.

2. Gli studi condotti dal gruppo di ricerca

Le nostre ricerche si inseriscono all’interno dell’ampio dibattito epistemologico e didattico su come integrare i risultati, i metodi, i quadri teorici e i risultati delle prove su larga scala - che sono progettati per avere un impatto a livello di sistema - su scala locale, a livello di scuole e insegnanti. Gli studi che presenteremo si riferiscono al contesto italiano. In Italia, le valutazioni su larga scala nazionali sono gestite dall’INVALSI (www.invalsi.it), l’ente di ricerca che, tra le varie mansioni, effettua

verifiche periodiche e sistematiche sulle conoscenze e abilità degli studenti e sulla qualità complessiva dell'offerta formativa delle istituzioni scolastiche. La consistenza (Looney, 2011) del quadro teorico delle prove INVALSI di matematica con le Indicazioni Curricolari Ministeriali¹ e i principali risultati di ricerca in didattica della matematica, come anche il disegno delle rilevazioni e le modalità di restituzione dei dati (INVALSI, 2018), fanno sì che i macro-fenomeni evidenziati in sede di rilevazione nazionale possano fornire informazioni utili e diventare strumenti interpretativi di alcuni aspetti dei processi di insegnamento e apprendimento della matematica specifici del contesto italiano.

Analisi qualitative e quantitative sui testi delle prove e sui risultati delle rilevazioni INVALSI hanno portato molti contributi nella ricerca in didattica della matematica da parte di diversi ricercatori italiani². Senza voler presentare in questo seminario un elenco esaustivo delle ricerche condotte dal nostro gruppo di ricerca, proponiamo una presentazione di alcune di esse che in qualche modo può delineare la mappa concettuale della nostra riflessione. I risultati ottenuti si collegano dialetticamente ad alcuni dei più importanti e condivisi risultati di ricerca in didattica della matematica, e al tempo stesso portano all'attenzione della comunità dei ricercatori nuovi fenomeni e nuove caratteristiche di fenomeni già conosciuti.

L'approccio di ricerca che accomuna i nostri studi è l'uso delle valutazioni su larga scala come strumento di riflessione, analisi e ricerca sui processi di apprendimento e insegnamento degli studenti italiani.

Decliniamo ora brevemente le componenti dell'approccio di ricerca che sta alla base delle ricerche che saranno oggetto del seminario.

Il primo principio comune è una visione fenomenologica dell'interazione del ricercatore coi processi di insegnamento-apprendimento; nei nostri studi partiamo da una osservazione ecologica di quello che viene percepito e rilevato sul campo dai diversi attori del sistema (ricercatori, insegnanti, allievi, INVALSI, ...), tenendo presente che comunque questa percezione e questa rilevazione sono a loro volta inquadrare e condizionate da convinzioni, teorie etc. Ma è proprio questa dialettica tra fenomenologia e teorie che permette l'evoluzione.

Tutte le nostre ricerche si basano inoltre su una visione formativa della valutazione in matematica abbracciando la definizione di valutazione formativa delineata all'interno del Progetto Europeo *LLP-Comenius FAMT&L* (Ferretti, Michael-Chrysanthou, & Vannini, 2018). Le valutazioni di classe e le valutazioni su larga scala hanno obiettivi e scopi diversi. Le valutazioni su larga scala sono valutazioni di sistema e hanno come obiettivo il monitoraggio di un sistema educativo e la verifica dell'implementazione di programmi ministeriali. Le informazioni restituite da queste valutazioni, quando coerenti con il sistema educativo, possono integrarsi nelle prassi scolastiche in un'ottica formativa (Looney, 2011; Bolondi, Ferretti, & Spagnuolo, 2016). La coerenza con il contesto educativo in cui sono inseriti i processi di insegnamento e apprendimento oggetto delle nostre ricerche è un fatto essenziale, come anche ritrovare nelle prove INVALSI esempi di quesiti in linea con gli obiettivi dichiarati nelle Indicazioni Ministeriali. Infatti, il quadro di riferimento delle prove INVALSI di matematica si richiama esplicitamente alle Indicazioni Ministeriali e agli altri riferimenti normativi e articola inoltre l'organizzazione delle prove sull'articolazione delle Indicazioni stesse (INVALSI, 2018). Questo fatto è cruciale e spiega il perché i nostri studi si possano basare solo sulle valutazioni INVALSI e non su qualsiasi altro tipo di valutazione su larga scala. Questa dichiarata

¹ Indicazioni Nazionali per il curriculum della scuola dell'infanzia e del primo ciclo d'istruzione https://www.miur.gov.it/documents/20182/51310/DM+254_2012.pdf/1f967360-0ca6-48fb-95e9-c15d49f18831?version=1.0&t=1480418494262, Assi Culturali

https://archivio.pubblica.istruzione.it/normativa/2007/allegati/all1_dm139new.pdf, Indicazioni Nazionali per i Licei https://www.indire.it/lucabas/lkmw_file/licei2010/indicazioni_nuovo_impaginato/Decreto_Indicazioni_Nazionali.pdf, Linee Guida per Istituti Tecnici e Professionali https://www.indire.it/lucabas/lkmw_file/nuovi_tecnici/INDIC/LINEE_GUIDA_TECNICI.pdf <https://nuoviprofessionali.indire.it/linee-guida-prof/>

² Ci sono naturalmente molti altri prodotti di ricerca curati o dal nostro gruppo di ricerca in collaborazione con altri ricercatori e insegnanti-ricercatori o da ricercatori di altri gruppi di ricerca.

coerenza delle prove con il curriculum *intended* va peraltro poi affiancata da una analisi della coerenza delle prove stesse con le pratiche effettive, che fanno parte del curriculum *implemented*, presenti nelle scuole italiane. Nelle ricerche si è sempre considerato come elemento essenziale sia della progettazione dei piani sperimentali, sia dell'interpretazione dei risultati, la rete di relazioni tra questi tre elementi: le Indicazioni nazionali, le prove effettivamente mandate sul campo dall'INVALSI, le pratiche dichiarate o rilevate nel contesto.

Dal punto di vista metodologico, nel senso di tecniche di raccolta e interpretazione di dati supportati dal sistema di principi, i nostri studi hanno utilizzato *mixed methods*, inserendo i materiali delle rilevazioni INVALSI in disegni sia di tipo *explanatory* sia di tipo *exploratory* (Creswell, & Plano-Clark, 2017). Una mixed-method research consiste essenzialmente in due o più componenti o trend di ricerca (Teddlie, & Tashakkori, 2009, Schoonenboom, & Johnson, 2017) provenienti da prospettive metodologiche diverse - nella maggior parte dei casi qualitative e quantitative. Si possono *combinare* le componenti tra loro, se l'obiettivo della ricerca è quello di unire e legare reciprocamente i risultati di ciascuna o *integrare* le componenti tra loro, se l'obiettivo è la convalida reciproca dei risultati della ricerca.

In tutti i nostri studi, la prima fase consiste nell'*integrazione* delle due componenti. Infatti, le ricerche che presenteremo nascono da una prima fase comune di individuazione di quelli che abbiamo definito *macro-fenomeni* emersi in sede di valutazione INVALSI.

Il macro-fenomeno è un fatto complesso, emergente dall'integrazione di due componenti, secondo Johnson and Onwuegbuzie (2004), una qualitativa e una quantitativa, integrate e complementari. Le informazioni quantitative provengono, dalle valutazioni INVALSI e questo è uno degli elementi fortemente caratterizzanti le nostre ricerche: le informazioni della componente quantitativa del macro-fenomeno sono ottenute da un quadro di rilevazione designato e effettuato da Ente di ricerca nazionale. Da qui la necessità, nei nostri studi, della condivisione del quadro teorico su cui sono costruiti i quesiti, le modalità di raccolta dei dati, le analisi effettuate e le modalità di restituzione dei dati, sia a livello campionario nazionale, sia a livello locale alle singole istituzioni scolastiche. La fase qualitativa del macro-fenomeno è condotta da ricercatori esperti di didattica della matematica, che con le lenti interpretative fornite da alcuni dei più affermati e condivisi risultati della ricerca analizzano e interpretano i risultati INVALSI delineando il macro-fenomeno.

A partire da questa prima fase comune, ciascuna ricerca è caratterizzata da specifiche domande di ricerca e da conseguenti impianti metodologici finalizzati alla loro risposta.

Le altre fasi delle ricerche a volte sono di tipo quantitativo, con l'utilizzo di dati provenienti da impianti sperimentali ad hoc, a volte di tipo qualitativo, e a volte di tipo mixed-method.

Durante il seminario, presenteremo alcuni degli studi che, in linea con questo approccio di ricerca, hanno permesso una caratterizzazione, quantificazione e misurazione della portata di diversi macro-fenomeni (come, ad esempio, nelle ricerche presentate in Ferretti, & Giberti, 2021, Ferretti, Giberti, & Lemmo, 2018, Ferretti, Santi, & Bolondi, 2022) o hanno messo in luce nuove evidenze e interpretazioni legate ai fenomeni stessi (fenomeni nuovi come l'effetto "Età della Terra", Ferretti, & Bolondi, 2019). Il gruppo di ricerca si è occupato anche dello studio di come questi macro-fenomeni possono costituire un ingrediente della formazione degli insegnanti, con quali strumenti e con quali modelli. Gli esempi di ricerche con cui concluderemo il seminario riguardano precisamente l'utilizzo di uno strumento di documentazione sulle prove nato per la formazione degli insegnanti, e un modello di formazione basato sull'analisi e discussione delle prove e dei loro risultati.

Nel corso del seminario verrà inoltre presentata la ricerca illustrata nel seguente paragrafo in cui si sono studiati alcuni aspetti predittivi delle prove INVALSI del grado 13 in relazione alla carriera universitaria (Gambini, De Simoni, & Ferretti, 2022); un esempio di ricerca che valida la coerenza delle prove con il contesto educativo italiano, autorizzando quindi a considerarne i materiali come una informazione utilizzabile criticamente per ricerche sulla scuola italiana.

2.1 Il potere predittivo delle prove INVALSI di matematica per il successo in ambito accademico

La validità e la coerenza delle prove INVALSI nel sistema educativo italiano è messa in luce anche dai risultati di una ricerca, di durata triennale, che mette in relazione i risultati dei pre-test INVALSI di matematica di Grado 13, i voti degli Esami di Stato conclusivi del secondo ciclo d'istruzione e i progressi in ambito matematico della carriera universitaria di studenti dell'Università di Bologna. Come mostrato dalla ricerca internazionale (Di Martino, Gregorio, & Iannone, 2022), la transizione tra la scuola secondaria e l'università è una questione problematica per diversi studenti, e, in particolare, la matematica gioca un ruolo cruciale sotto diversi punti di vista. Ciò accade anche nel contesto italiano (si veda, ad esempio, Di Martino, & Gregorio, 2019). Recenti ricerche (Camara, 2013) hanno messo in luce il problema dell'identificazione di indici attendibili e validi di preparazione degli studenti della scuola secondaria in vista dell'accesso all'università, nonché la necessità di validi strumenti per prevedere il successo accademico degli studenti del primo anno. Il nostro studio (Gambini, De Simoni, & Ferretti, 2022) si inserisce in questa linea di ricerca e prende in esame le relazioni tra i risultati ottenuti nei pre-test INVALSI di matematica per il grado 13 dell'a.s. 2017-18, i voti finali dell'esame di Stato e il successo in matematica in riferimento alla carriera universitaria dei primi due anni di 177 laureandi dell'Università di Bologna. In linea con la letteratura, i nostri risultati confermano il ruolo dei voti dell'Esame di Stato come predittori del "successo accademico" nell'istruzione universitaria anche nel contesto italiano. La ricerca mostra anche che i risultati ottenuti nei pre-test INVALSI migliorano il potere predittivo dei voti dell'Esame di Stato non solo ai fini della previsione del successo nei corsi di matematica del primo anno, ma anche quando si tratta di criteri più generali di successo universitario. Ad esempio, con poche eccezioni, i nostri risultati mostrano che i risultati ottenuti nei pre-test INVALSI abbiano un potere predittivo per quanto riguarda il superamento e i risultati ottenuti negli esami di matematica del primo anno più forte rispetto a quello dei voti dell'Esame di Stato. Lo studio svolto è esplorativo ed ha alcune limitazioni dovute al campione, all'analisi di solo quattro corsi universitari e al fatto che i pre-test INVALSI considerati fossero quelli del primo anno del disegno della rilevazione per il grado 13, ancora in una fase di assestamento. Anche in considerazione di queste evidenze, il nostro approccio si basa su un utilizzo ragionato dei dati statistici e psicometrici, che tenga conto delle specifiche caratteristiche di questi dati.

3. Le analisi statistiche su cui si basano alcune delle nostre ricerche

Le ricerche presentate in questo Seminario Nazionale prendono in considerazione il quadro teorico e metodologico proposto dall'Istituto INVALSI nei due principali documenti relativi alle prove: il *Rapporto dei Risultati* e il *Rapporto Tecnico*. Le metodologie quantitative di analisi dei risultati delle prove utilizzate nelle ricerche sono pertanto pienamente coerenti con le metodologie adottate dall'Istituto e la scelta degli indici e dei modelli statistici utilizzati viene ampiamente illustrata all'interno del Rapporto Tecnico (si veda, ad esempio, INVALSI, 2017). Vi sono infatti due principali approcci che possono essere adottati per analizzare rilevazioni su larga scala: la Teoria Classica dei test (CTT) e la Teoria della Risposta all'Item (IRT) (Barbaranelli, & Natali, 2005) ed entrambi vengono presi in considerazione per l'analisi dei dati raccolti dalle prove INVALSI all'interno del rapporto tecnico. La CTT fornisce importanti strumenti statistici per lo studio delle prove e permette una prima analisi relativa alla difficoltà dei quesiti o degli item che li compongono (intesa come percentuale di risposte corrette), discriminatività degli item (ovvero capacità di ogni item di distinguere studenti con livelli diversi di abilità) e coerenza interna della prova, in termini di Alpha di Cronbach (Barbaranelli, & Natali, 2005). Nonostante ciò, le analisi e le restituzioni dei risultati delle prove INVALSI sono basate principalmente sulla più moderna IRT (Rasch, 1960) che permette di superare le principali limitazioni della CTT tra cui, per esempio, la dipendenza tra la stima dell'abilità dei soggetti e la difficoltà dei quesiti della prova. L'interpretazione qualitativa dei risultati, in chiave didattica e di formazione degli insegnanti, può avvalersi proficuamente di una integrazione dei due approcci statistici.

Il modello di Rasch è il più semplice tra i modelli afferenti alla IRT, si tratta infatti di un modello logistico a un parametro che opera una stima congiunta tra l'indice di difficoltà relativo a ogni quesito della prova e l'indice di abilità attribuito a ogni studente.

Nello specifico, tale modello permette di esprimere la probabilità di fornire la risposta corretta a un quesito di una prova in funzione della difficoltà del quesito stesso e dell'abilità dello studente misurata sull'intera prova. Questa relazione tra l'abilità del soggetto e la probabilità di rispondere correttamente al quesito può essere rappresentata graficamente attraverso la curva teorica ipotizzata dal modello, detta anche curva caratteristica dell'item (Item Characteristic Curve – ICC).

Una volta applicato il modello di Rasch, osservando specifici grafici, chiamati *distractor plot* (Figura 1), possono essere estrapolate diverse informazioni. Nello stesso grafico in cui viene rappresentata la curva teorica ICC relativa alla risposta corretta (rappresentata con una linea continua), vengono rappresentati anche i dati empirici relativi alla risposta corretta e alle altre opzioni di risposta (linee spezzate).

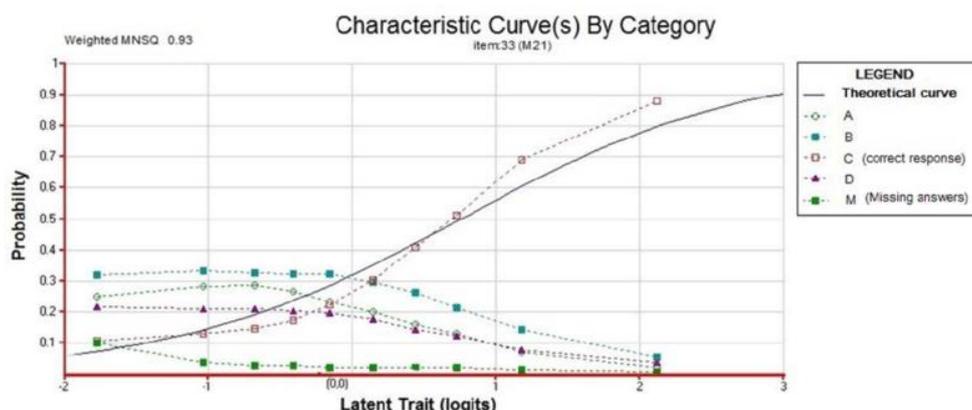


Figura 1. Esempio di curva caratteristica di un quesito INVALSI a risposta multipla

In questo modo è possibile osservare quanto la curva empirica della risposta corretta sia coerente con la curva teorica e, inoltre, si può analizzare l'andamento delle curve empiriche di ogni distrattore (inteso come risposta non corretta) mettendo in relazione la probabilità di scelta di quell'opzione di risposta (rappresentata sull'asse y) con il livello di abilità degli studenti – misurata in riferimento alla prova (rappresentata sull'asse x).

4. Un primo esempio: macro-fenomeni emersi da quesiti con curve empiriche dei distrattori con andamento “a pancia”

In generale, al crescere dell'abilità dello studente nella prova, si dovrebbe osservare una probabilità maggiore che egli fornisca una risposta corretta mentre le opzioni di risposta errate, le curve riferite ai distrattori, dovrebbero mostrare un andamento decrescente al crescere dell'abilità degli studenti. In figura si può notare che la curva caratteristica e la curva empirica dell'opzione corretta sono crescenti mentre le curve empiriche dei distrattori sono decrescenti. Da un punto di vista strettamente statistico ci si aspetta quindi che, in un item, al crescere dell'abilità degli studenti cresca la percentuale di risposte corrette e, al contempo, diminuisca la percentuale di risposte sbagliate.

Osservando i quesiti a risposta multipla, in cui la risposta corretta è unica ma sono proposte altre due/tre opzioni di risposta errate, si dovrebbe notare che la percentuale di risposte errate (considerate nel loro complesso e tenendo conto anche delle risposte mancanti) è decrescente al crescere del livello di abilità degli studenti nella prova. Esistono però casi particolari in cui le curve empiriche relative ai singoli distrattori mostrano andamenti non strettamente decrescenti (Figura 2), abbiamo denominato andamento “a pancia” questo particolare andamento, non strettamente decrescente.

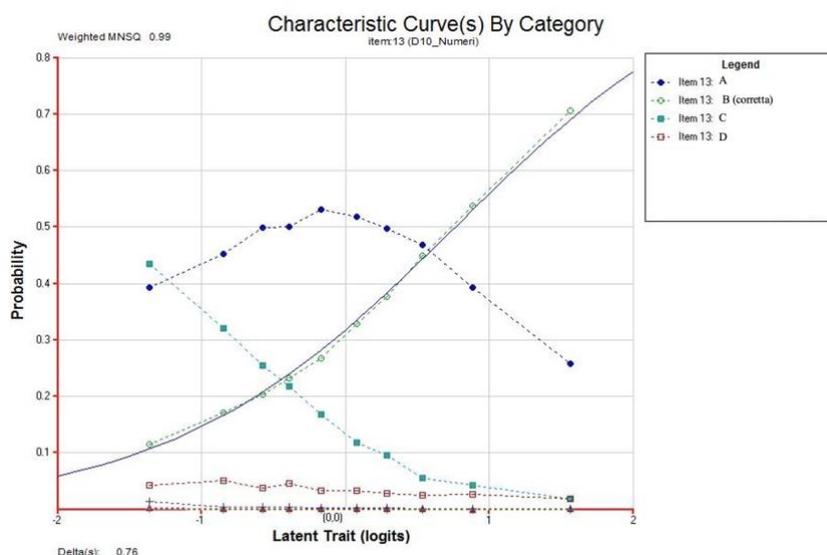


Figura 2. Curva caratteristica di un quesito con la curva di un distrattore con andamento “a pancia”

Le opzioni di risposta che mostrano questo particolare andamento “a pancia” sono opzioni errate che risultano particolarmente attrattive per studenti con livelli di abilità medi o medio-alti nella prova. Ed è proprio l’analisi dell’andamento di queste curve che sarà alla base di alcune delle ricerche che illustreremo. In particolare, i quesiti presi in considerazione in queste ricerche sono tratti da prove INVALSI di diversi livelli scolastici (dalla scuola primaria alla scuola secondaria di secondo grado), presentano macro-fenomeno dal punto di vista didattico, mostrano buone proprietà misuratorie (INVALSI, 2017; Barbaranelli, & Natali, 2005). L’analisi di questi quesiti è stata alla base di diverse ricerche (Bolondi, Ferretti, & Giberti, 2018; Ferretti, Giberti, & Lemmo, 2018; Ferretti, & Giberti, 2021; Giberti, 2018) condotte con l’obiettivo di indagare caratteristiche e peculiarità dei macro-fenomeni emersi dai quesiti le cui i distractor plot presentano questo andamento. I primi studi (Ferretti, Giberti, & Lemmo, 2018) hanno evidenziato quanto i macro-fenomeni emersi da diversi di questi quesiti, fossero manifestazioni di fenomeni inquadrabili con alcuni dei costrutti più noti e diffusi in letteratura, come quello di contratto didattico nel senso di Brousseau (1997). Gli studi hanno così permesso non solo una conferma e una quantificazione nel contesto educativo italiano di quanto emerso in letteratura, ma hanno permesso di studiare a fondo il fenomeno fornendone una caratterizzazione in termini di studenti che hanno mostrato di avere raggiunto determinate performance in riferimento alle prove svolte.

Ulteriori sviluppi, hanno permesso di caratterizzare alcuni dei macro-fenomeni emersi anche in termini di differenze di genere e i risultati ottenuti offrono spunti di riflessioni per quanto riguarda il legame tra il contratto didattico e il gender gap (Bolondi, Ferretti, & Giberti, 2018; Ferretti, & Giberti, 2021).

Come vedremo, le analisi di due quesiti aventi curve caratteristiche con questo andamento “a pancia” sono anche alla base della ricerca che ha permesso di mettere a punto un nuovo effetto di contratto didattico, l’effetto “età della terra”.

5. L’effetto “Età della terra”

Il punto di partenza della ricerca³ che ha condotto alla definizione dell’effetto è un fenomeno didattico che è stato osservato in due quesiti delle prove INVALSI di matematica, effettuate in Italia da tutti gli studenti di grado 10 (secondo anno della scuola secondaria di II grado). Il primo quesito,

³ La ricerca è oggetto della tesi di dottorato di Federica Ferretti reperibile al seguente link: <http://amsdottorato.unibo.it/7213/#>

somministrato in modo censuario a circa 600 mila studenti italiani di Grado 10 di tutti i percorsi scolastici (Licei, Istituti Tecnici e Istituti Professionali) è rappresentato in Figura 3.

D5. L'età della Terra è valutata intorno ai $4,5 \times 10^9$ anni. L'Homo Erectus è comparso circa 10^6 anni fa. Qual è la stima che più si avvicina all'età che la Terra aveva quando è comparso l'Homo Erectus?

A. $4,5 \times 10^9$ anni

B. $3,5 \times 10^9$ anni

C. $4,5 \times 10^6$ anni

D. $4,5 \times 10^3$ anni

Figura 3 Quesito D5, Grado 10 – Prova INVALSI di matematica 2011

Nel grafico seguente (Figura 4) sono riportati i risultati in riferimento a un campione statistico rappresentativo a livello nazionale di circa 45.000 studenti.

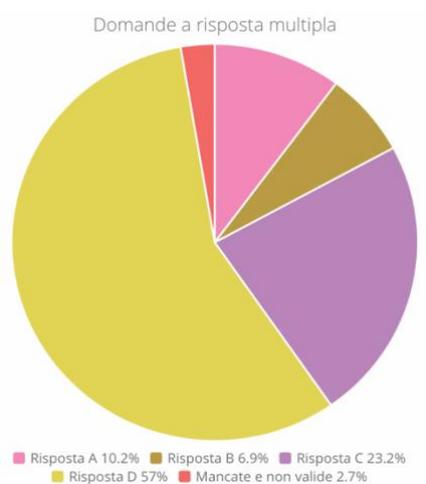


Figura 4. Percentuali Campione nazionale, Quesito D5, Grado 10 – Prova INVALSI di matematica 2011, www.gestinv.it

Come si vede dal grafico (Figura 4), le risposte corrette (opzione A) sono state il 10,2%. L'opzione sbagliata più scelta è l'opzione D, in cui l'esponente della potenza è la differenza degli esponenti delle potenze nel testo. Una situazione equivalente, dal punto di vista matematico, si è presentata nella prova INVALSI 2013, sempre nella prova per il grado 10. Il quesito (Figura 5) è stato somministrato a 560.000 studenti delle classi seconde del secondo ciclo di ogni percorso scolastico e il campione era costituito da circa 40.000 studenti.

D6. Un atomo di idrogeno contiene un protone la cui massa m_p è all'incirca $2 \cdot 10^{-27}$ kg, e un elettrone la cui massa m_e è all'incirca $9 \cdot 10^{-31}$ kg. Quale tra i seguenti valori approssima meglio la massa totale dell'atomo di idrogeno (cioè $m_p + m_e$)?

A. $2 \cdot 10^{-27}$ kg

B. $11 \cdot 10^{-31}$ kg

C. $11 \cdot 10^{-58}$ kg

D. $18 \cdot 10^{-58}$ kg

Figura 5. Quesito D6, Grado 10 – Prova INVALSI di matematica 2013

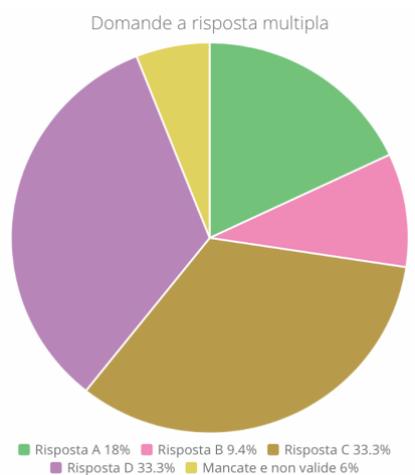


Figura 6. Risultati – Campione nazionale, Quesito D6, Grado 10 – Prova INVALSI di matematica 2013, www.gestinv.it

Come mostrano i risultati (Figura 6), le risposte corrette sono state il 18%. Analizzando le percentuali di scelta delle opzioni sbagliate, si può vedere come le opzioni C e D, in cui l'esponente della potenza è ottenuto per somma degli esponenti delle potenze presenti nel testo, sono i più scelti a livello nazionale (più del 30% di scelte ciascuno).

Le prime interpretazioni (cfr. ad es. Impedovo, Orlandoni, & Paola, 2011) del fenomeno evidenziato dalla domanda “età della Terra” hanno collegato il comportamento degli allievi genericamente a effetti di contratto didattico nel senso di Brousseau (Brousseau, 1997) e alla mancanza di controllo critico sui contenuti. In particolare, il fatto che la maggior parte degli allievi nel quesito “età della Terra”, il cui testo richiama intuitivamente una situazione sottrattiva, scelgano l'opzione D in cui l'esponente della potenza è la differenza degli esponenti delle potenze presenti nella consegna potrebbe apparire come un evidente esempio di presenza della clausola del contratto didattico denominata delega formale (D'Amore, 2007); così come nel quesito “Massa del protone” (situazione additiva) la maggior parte degli studenti scelgono le opzioni C e D (in cui compare la somma degli esponenti delle potenze),

Osserviamo però un altro fatto, manifestato in entrambi i casi: la risposta corretta è uno dei dati esplicitamente presenti nel testo del quesito. In due quesiti somministrati agli studenti del grado 8 (classe terza della scuola secondaria di primo grado) si trovano situazioni simili (quesito E21, PN a.s. 2011/12 e quesito D23, PN a.s. 2009/10). Comune a questi quattro esempi, differenti per molti aspetti di formulazione, contesto, contenuto, abilità richiesta, è la difficoltà che si manifesta quando la risposta coincide con uno dei dati di partenza, difficoltà che coinvolge inaspettatamente anche allievi di livello medio e alto di abilità nella prova. Da queste osservazioni, è nata la ricerca che portano alla definizione di un nuovo effetto di contratto didattico denominato effetto “età della Terra”. In particolare, all'interno della ricerca sono stati indagati, e definiti, l'effetto e il principio regolativo che lo determina. In dettaglio, l'effetto che si osserva può essere descritto in questo modo:

In una situazione didattica, di fronte a una consegna gli allievi tendono a non accettare una risposta che non sia identificabile chiaramente con un risultato distinto dai dati di partenza.

Ipotesi di ricerca è che questo effetto dipenda da un principio regolativo (che a una prima analisi sembra discendere da pratiche d'aula) del tipo:

Il risultato di un problema o di una operazione non può essere uguale al dato di partenza.

Nell'ambito della ricerca si è indagato se il fenomeno si manifesta in ambiti matematici diversi, se interessa tutti i livelli scolastici, se è presente in studenti con abilità matematiche diverse, se dipende dal controllo semantico che l'allievo ha sui contenuti della domanda e dalle diverse abilità matematiche e, infine, se è possibile quantificarlo. In secondo luogo, si è indagato se il comportamento degli allievi in cui si osserva l'effetto dipende da un principio regolativo vicino a quello enunciato. Per rispondere alle domande di ricerca è stata condotta una sperimentazione,

costituita da fasi quantitative e qualitative (Johnson, & Onwuegbuzie, 2004), che ha coinvolto circa 650 studenti del primo e del secondo ciclo d'istruzione. I risultati mostrano che l'effetto si manifesta in tutti i livelli scolastici indagati e non dipende né dal controllo semantico della situazione proposta né dall'abilità matematica in riferimento al contenuto indagato. L'ampliamento del quadro teorico di partenza ha permesso di inquadrare le interviste condotte dagli studenti e, quindi, il principio regolativo determinante l'effetto. Dalla ricerca emerge che l'effetto si manifesta in minor misura per quanto riguarda quesiti di ambito geometrico. I principali risultati della ricerca sono stati oggetto di pubblicazione in Ferretti e Bolondi (2019).

6. Prove INVALSI e formazione degli insegnanti

Quanto messo in luce dai macro-fenomeni emersi in sede di valutazione standardizzata, se opportunamente e consapevolmente integrato in ambito di formazione insegnanti, può contribuire a migliorare conoscenze e abilità degli insegnanti (Campbell, & Levin, 2009; Di Martino, & Baccaglini-Frank, 2019; Ferretti, Lemmo, & Martignone, 2018b; Martignone, 2017). L'analisi dei risultati di queste prove standardizzate sul campione nazionale e per le diverse scuole, interpretati alla luce dei singoli contesti, può essere utile agli insegnanti per riflettere sui processi di insegnamento-apprendimento e anche sui possibili punti di forza e criticità di specifiche scelte didattiche.

Per questo, già da molti anni, diversi membri del gruppo di ricerca hanno utilizzato le prove INVALSI in ambito di formazione insegnanti di matematica in formazione e in servizio (Bolondi, Ferretti, & Spagnuolo, 2016; Bolondi et al., 2016; Lemmo et al., 2015; Martignone, 2016; Ferretti, Lemmo, & Martignone, 2018a).

Con il passare del tempo e con l'aumento dell'esperienza sia in campo di formazione sia in ambito di ricerca sul legame tra la didattica della matematica e le valutazioni su larga scala, le esperienze nel campo della formazione insegnanti che utilizzavano le prove INVALSI sono confluite in ricerche sullo sviluppo professionale degli insegnanti e dei futuri insegnanti di matematica.

In dettaglio, le attività che abbiamo sviluppato e studiato si basano su un utilizzo consapevole delle prove INVALSI di matematica per contribuire allo sviluppo delle conoscenze specialistiche degli insegnanti e dei futuri insegnanti. Nella sezione successiva sarà illustrato un modello di formazione insegnanti elaborato negli ultimi anni che collega l'analisi delle prove INVALSI di matematica e dei risultati di queste con lo sviluppo professionale degli insegnanti in formazione e in servizio innescando un ciclo virtuoso tra valutazione standardizzata, formazione insegnanti e insegnamento della matematica. Durante il seminario saranno discussi degli esempi di implementazione del modello con insegnanti in formazione e in servizio. Il seminario si concluderà con riflessioni in riferimento ad alcuni risultati di un progetto di ricerca condotto dal gruppo di ricerca *dell'Osservatorio SIRD – INVALSI e didattiche disciplinari* (Arzarello, & Ferretti, 2021). Il progetto di ricerca è attualmente volto a raccogliere e analizzare informazioni sulle conoscenze e gli atteggiamenti degli insegnanti nei confronti dell'INVALSI: sia dell'INVALSI come Istituto, con le sue finalità e modalità di lavoro, sia nei confronti delle prove INVALSI di matematica e delle loro ricadute sulla didattica.

6.1 Un modello di formazione per insegnanti e futuri insegnanti di matematica che utilizza le prove INVALSI

I nostri studi sulla formazione insegnanti sono incentrati da un lato su una prospettiva teorica per la progettazione di metodologie efficaci, dall'altro sull'introduzione di strumenti e risorse utilizzati e progettati per una collaborazione efficace tra insegnanti e tra insegnanti e ricercatori/formatori. La progettazione del nostro modello può essere inquadrata negli studi sulle *Community of inquiry* (Jaworski, 2006) e utilizza i risultati delle ricerche legate al modello *Mathematics Teacher's Specialised Knowledge* (MTSK) di Carrillo-Yañez e colleghi (2018). Riteniamo che le lenti teoriche

offerte da questi studi permettano di inquadrare e definire due peculiarità distintive caratterizzanti del modello di formazione degli insegnanti che abbiamo elaborato. Nelle nostre ricerche volevamo concettualizzare e delineare le conoscenze e le competenze professionali specifiche che gli insegnanti potrebbero raggiungere come risultato del cambiamento (Guskey, 2002) innescato nel loro processo di formazione. Ci riferiamo a un'ampia gamma di conoscenze che includono matematica, epistemologia, pedagogia, didattica disciplinare, psicologia, ecc. i nostri studi si sono quindi avvalsi del modello MTSK per individuare e analizzare le conoscenze specialistiche di matematica degli insegnanti. All'interno di una *Community of inquiry* lo sviluppo di conoscenze specialistiche può richiedere strumenti e, in senso più ampio, risorse che mediano l'attività, contribuiscono allo scambio interpersonale tra i suoi attori e fanno emergere oggetti culturali e concettuali. Come sostenuto da Adler (2000), la formazione degli insegnanti di matematica dovrebbe incentrarsi sulle risorse, intese come strumenti che consentono agli insegnanti di ampliare il loro impatto nell'attività matematica in classe. In questa linea, i nostri studi mettono in gioco uno strumento/risorsa introdotto nel 2014 contenente tutti i quesiti delle valutazioni standardizzate INVALSI (Bolondi, Ferretti, & Gambini, 2017): il database delle prove INVALSI *Gestinv* (www.gestinv.it).

Gestinv è un database strutturato contenente tutti i dati delle valutazioni standardizzate INVALSI dalle rilevazioni del 2008 di tutti gli ambiti oggetto di indagine. In dettaglio, contiene 2340 item delle prove INVALSI di matematica; l'impatto del database è stato valutato sia quantitativamente che qualitativamente, attraverso indicatori standard come il numero di utenti registrati (oltre 33.000), il numero di accessi (in media 200 ogni giorno), il tempo trascorso sul sito e altri parametri. Questi dati, insieme alle sue informazioni strutturate, promuovono Gestinv come uno strumento da implementare nella progettazione di modelli di formazione degli insegnanti (Ferretti, Gambini, & Santi, 2020).

Ci sono molti modi in cui si può utilizzare il database; al suo interno si possono infatti effettuare diverse forme di ricerca. In Gestinv ci sono i quesiti relativi alle prove INVALSI di tutte le materie oggetto di rilevazione, italiano, matematica e inglese. Noi presenteremo la sezione riguardante le prove INVALSI di matematica, l'unica oggetto del nostro modello di formazione. Nella nostra ricerca, Gestinv svolge un ruolo importante nel fornire agli insegnanti e ai ricercatori uno strumento interattivo per accedere a una vasta gamma di informazioni riguardanti le prove e i risultati delle prove INVALSI sul campione nazionale (Bolondi, Ferretti, & Gambini, 2017, Ferretti, Gambini, & Santi, 2020). In particolare, come abbiamo visto, i risultati delle prove INVALSI evidenziano e quantificano macro-fenomeni rilevanti, che possono essere interpretati secondo i metodi e i risultati della ricerca in didattica della matematica. La struttura articolata e la ricchezza di informazioni fornite da Gestinv collega la pratica svolta in classe con la conoscenza specialistica degli insegnanti di matematica. Riteniamo che Gestinv sia una risorsa, nel senso sopra indicato, che intreccia strumenti "concettuali" e "pratici" (Grossman, Smagorinsky, & Valencia, 1999). Gestinv è uno strumento che può essere implementato, con un'utilità locale e immediata, come fonte di materiale didattico per l'insegnamento e l'apprendimento della matematica e per questo possiamo considerarlo uno strumento "pratico".

D'altra parte, il quadro teorico (INVALSI, 2018) che informa sia la costruzione che la selezione dei quesiti e la complessa struttura per la classificazione e ricerca delle informazioni sui quesiti nel database di Gestinv possono innescare l'uso di conoscenze di contenuto matematico, di strumenti teorici dell'educazione matematica, di idee sull'insegnamento e sull'apprendimento che complessivamente costituiscono strumenti concettuali.

6.2 Esempio di ricerca effettuata con Gestinv: un macro-fenomeno emerso in sede INVALSI in ambito algebrico

La ricerca che presentiamo (oggetto di pubblicazione in Ferretti, Santi, & Bolondi, 2022) è volta ad indagare la padronanza di alcuni aspetti matematici in ambito algebrico da parte degli studenti della scuola secondaria di secondo grado e, come vedremo, è un esempio di risultato di ricerca che si presta ad essere utilizzata nelle prime fasi del nostro modello di formazione. Per indagare le difficoltà degli

studenti italiani in ambito algebrico abbiamo utilizzato come strumento di ricerca il database Gestinv (www.gestinv.it). Il database Gestinv ci ha permesso di effettuare un'analisi quantitativa basata sui quesiti INVALSI di algebra di grado 10 che hanno avuto percentuali di risposta corrette molto basse a livello nazionale e i cui contenuti fossero vicini a quelli normalmente affrontati nelle scuole superiori italiane durante i primi due anni di scuola secondaria.

Tra tutti i quesiti restituiti dal database, sono risultati particolarmente interessanti due quesiti riguardanti le operazioni di trattamento (in senso di Duval, 2006) con disuguaglianze. In particolare, i due quesiti sono caratterizzati dal fatto che la soluzione non richiede necessariamente operazioni di trattamento esplicite. A partire dal macro-fenomeno emerso abbiamo condotto uno studio qualitativo per inquadrare le possibili strategie messe in atto dagli studenti italiani e per avere un quadro interpretativo di risultati così diffusamente fallimentari.

Come vedremo dettagliatamente, per analizzare l'apprendimento dell'algebra abbiamo utilizzato la Teoria della Reificazione di Sfard (1991). Questa teoria considera i concetti matematici caratterizzati da una doppia natura di processo e oggetto: sono due aspetti complementari dell'oggetto matematico. Per avere una lente teorica che permetta di caratterizzare il lato-processo e il lato-oggetto abbiamo utilizzato la Teoria dell'Oggettivazione di Radford (2006, 2021) e l'approccio strutturale di Duval (2006, 2017).

Lo studio, basato sull'analisi delle principali strategie per la risoluzione dei quesiti, mette in luce che gli studenti non hanno oggettivato, nel senso di Radford (2021), il significato corretto delle disuguaglianze. Il fragile controllo semantico nella comprensione delle disuguaglianze e il significato non radicato in un adeguato processo di oggettivazione si evincono anche dalle difficoltà nella gestione sia di trattamenti, che secondo gli studi Duval (2017) dovrebbero invece essere supportati dalla sintassi del registro algebrico, sia di conversioni che, in questi casi, renderebbero le soluzioni banali.

6.3 Il nostro modello di formazione

Come vedremo dettagliatamente durante il seminario, il modello che proponiamo per lo sviluppo professionale degli insegnanti in formazione e in servizio si basa sull'utilizzo del database Gestinv all'interno di una *Community of Inquiry* (Jaworski, 2006). I nostri studi riguardano l'influenza che l'uso consapevole di Gestinv può avere sulle conoscenze specialistiche degli insegnanti di matematica riflettendo criticamente sulla complessità delle valutazioni standardizzate con il supporto dei risultati della ricerca in educazione matematica. Come abbiamo già scritto, il quadro teorico con cui inquadrare le conoscenze specialistiche degli insegnanti e dei futuri insegnanti di matematica è il modello *Mathematics Teacher Specialized Knowledge* (MTSK) di Carrillo-Yañez e colleghi (2018). Ispirato dagli studi di Shulman (1986), il modello MTSK considera due ampie aree di quelle che considerano conoscenze specialistiche degli insegnanti di matematica: la Conoscenza Matematica (MK) e la Conoscenza Pedagogica del Contenuto (PCK).

I *beliefs* degli insegnanti sulla matematica e sul suo apprendimento e insegnamento si trovano al centro del modello MTSK proprio per sottolineare la reciprocità tra *beliefs* e domini di conoscenza identificati nel modello (Carrillo-Yañez et al., 2018). Come illustreremo dettagliatamente durante il seminario, riteniamo il modello MTSK idoneo ad inquadrare i diversi aspetti delle conoscenze specialistiche degli insegnanti da noi studiati.

La metodologia generale del nostro modello di formazione è costituita dalle seguenti fasi:

- Introduzione dell'attività. I ricercatori/formatori affrontano il contenuto matematico selezionato per l'attività da un punto di vista epistemologico e didattico. Successivamente vengono presentate alcune delle funzionalità di Gestinv che gli insegnanti useranno nella loro indagine e analizzano degli esempi di quesito. I ricercatori/formatori discutono con gli

insegnanti i principali costrutti di didattica della matematica che forniscono possibili chiavi di lettura dei macro-fenomeni che emergono.

- Analisi di un esempio. I ricercatori discutono con l'intero gruppo di insegnanti un macro-fenomeno didattico utilizzando Gestinv stimolando riflessioni su uno o più sottodomini della conoscenza specialistica descritta dal modello MTSK.
- Attività di gruppo. I ricercatori assegnano un compito agli insegnanti che lavoreranno suddivisi in sottogruppi di massimo 4/5 persone. Il compito coinvolge un contenuto matematico in riferimento alle Indicazioni ministeriali del grado scolastico degli insegnanti coinvolti ed è inerente a una o più difficoltà di apprendimento individuate in letteratura. L'attività di piccolo gruppo si svolge in linea con le caratteristiche di una *Community of inquiry*. L'attività di gruppo mira alla costruzione di un prodotto multimediale, un artefatto, la progettazione di un'attività per gli studenti, ecc. Partendo da questi elaborati si discuteranno le conoscenze e i *beliefs* degli insegnanti coinvolti.
- Discussione generale. I sottogruppi presentano le loro produzioni al grande gruppo e al ricercatore. Ogni presentazione viene discussa all'interno della community of inquiry al fine di evidenziare *beliefs*, affrontare dubbi, difficoltà e contenuti poco chiari riguardo sia la MK sia la PCK e delineare i sottodomini del MTSK che sono emersi durante tutto il percorso. La discussione finale, basata sulle presentazioni condivise, si svolge con le stesse caratteristiche di una *Community of inquiry*.

La valutazione risulta essere un elemento fondamentale della *identity* dell'insegnante (Ferretti, Del Zozzo, & Santi, 2020; Ferretti et al., 2021; Goldin et al., 2016) e la formazione effettuata su questo aspetto, supportata anche da strumenti come Gestinv, ha fornito ulteriori materiali su cui sono state sviluppate ricerche, che saranno presentate nel seminario (come, ad esempio, Ferretti, Martignone, & Santi, 2022).

7. Conclusioni

L'insieme delle ricerche presentate mostra, a nostro avviso, che il patrimonio di informazioni restituito sulla scuola italiana dalle prove INVALSI di matematica può anche costituire una risorsa preziosa per i ricercatori, una volta che questi risultati sono interpretati nel contesto specifico e con adeguati strumenti teorici. Al termine del seminario verranno anche presentate possibili nuove direzioni di ricerche.

Bibliografia

- Adler, J. (2000). Conceptualising resources as a theme for teacher education. *Journal of Mathematics Teacher Education*, 3(3), 205-224.
- Arzarello, F., & Ferretti, F. (2021). The connection between the mathematics INVALSI test and the teaching practices: an explorative study. *I dati INVALSI: uno strumento per la ricerca*. FrancoAngeli.
- Atkin, J. M. (1998). The OECD study of innovations in science, mathematics and technology education. *Journal of Curriculum Studies*, 30(6), 647-660.
- Barbaranelli, C., & Natali, E. (2005). I Test Psicologici. *Teorie e Modelli Psicometrici*. Roma: Carocci.
- Bolondi, G., Branchetti, L., Ferretti, F., Lemmo, A., Maffia, A., Martignone, F., Matteucci, M. L., Mignani, S. & Santi, G. R. P. (2016). Un approccio longitudinale per l'analisi delle prove INVALSI di matematica: Cosa si può dire degli studenti in difficoltà. In *Concorso di idee per la ricerca: progetto Sistema informativo integrato I-3-FSE-2009-1: PON FSE Competenze per lo sviluppo: Convenzione MIUR 24/04/2009* (pp. 81-102). Cleup.

- Bolondi, G., Ferretti, F., & Gambini, A. (2017). Il database Gestinv delle prove standardizzate INVALSI: Uno strumento per la ricerca. In P. Falzetti (Ed.), *I dati INVALSI: Uno strumento per la ricerca* (pp. 33–42). Milano: FrancoAngeli.
- Bolondi, G., Ferretti, F., & Giberti, C. (2018). Didactic contract as a key to interpreting gender differences in maths. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal)*, (18), 415-435.
- Bolondi G., Ferretti F., Spagnuolo A. (2016), “Le prove INVALSI con Geogebra: trasformare la valutazione standardizzata in valutazione formativa”, in Robutti O. (a cura di), *La formazione docenti con Geogebra. Atti del IV GeoGebra Italian Day 2014*, Ledizioni, Milano: 99-108.
- Bolondi, G., & Ferretti, F. (2021). Quantifying Solid Findings in Mathematics Education: Loss of Meaning for Algebraic Symbols. *International Journal of Innovation in Science and Mathematics Education*, 29(1), 1-15.
- Breakspear, S. (2012). The policy impact of PISA: an exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers, No. 71*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/5k9fdfqffr28-en>
- Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht: Kluwer Academic Publishers.
- Camara, W. (2013). Defining and measuring college and career readiness: A validation framework. *Educational Measurement: Issues and Practice*, 32(4), 16-27.
- Campbell, C., & Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 47-65.
- Carnoy, M. (2015). *International Test Score Comparisons and Educational Policy. A Review of the Critiques*. Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/files/pb_carnoy_international_test_scores_0.pdf
- Carrillo-Yañez, J., Climent, N., Montes, M., Contreras, L. C., Flores-Medrano, E., Escudero-Ávila, D., Vasco, D., Nielka, R., Pablo, F., Álvaro, A. G., Ribeiro, M., Muñoz-Catalán, M. C. (2018). The mathematics teacher’s specialised knowledge (MTSK) model. *Research in Mathematics Education*, 20(3), 236-253.
- Cochran-Smith, M. (2001). Learning to Teach Against the (New) Grain. *Journal of Teacher Education*, 52(1), 3–4.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). London, UK: Sage.
- D’Amore, B. (2007). Epistemologia, didattica della matematica e pratiche d’insegnamento. *La matematica e la sua didattica*, 21(3), 347-369.
- De Lange, J. (2007). Large-scale assessment and mathematics education. *Second handbook of research on mathematics teaching and learning*, 2, 1111-1144.
- Di Martino, P., & Baccaglioni-Frank, A. (2017). Beyond performance results: Analyzing the informational and developmental potentials of standardized mathematics tests. *For the Learning of Mathematics*, 37(3), 39-44.
- Di Martino, P., Gregorio, F., & Iannone, P. (2022). The transition from school to university in mathematics education research: new trends and ideas from a systematic literature review. *Educational Studies in Mathematics*, 1-28.
- Di Martino, P., & Gregorio, F. (2019). The mathematical crisis in secondary–tertiary transition. *International Journal of Science and Mathematics Education*, 17(4), 825-843.
- Doig, B. (2006). Large-scale mathematics assessment: looking globally to act locally. *Assessment in Education: Principles, Policy & Practice*, 13(3), 265–288.
- Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational studies in mathematics*, 61(1), 103-131.
- Duval, R. (2017). *Understanding the mathematical way of thinking: The registers of semiotic representations*. Springer International Publishing.

- Eurydice, I. U. I. (2016). La valutazione delle scuole in Europa: politiche e approcci in alcuni Paesi europei. *Firenze, Eurydice Italia*, 83.
- Ferretti, F., Bolondi, G. (2019). This cannot be the result! The didactic phenomenon “the Age of the Earth”. *International Journal of Mathematical Education in Science and Technology*. 52(2), 194-207.
- Ferretti, F., Del Zozzo, A., & Santi, G. (2020). La didattica della matematica a distanza ai tempi del Covid-19 e la sua interazione con l’identità docente. *Annali online della Didattica e della Formazione Docente*, 12(20), 84-108.
- Ferretti, F. & Gambini, A. (2017). A vertical analysis of difficulties in mathematics by secondary school to level; some evidences stems from standardized assessment. In T. Dooley & G. Gueudet (Eds.), *Proc. of the Tenth Congress of the European Society for Research in Mathematics Education* (pp. 3492- 3499). Dublin, Ireland: DCU Institute of Education and ERME.
- Ferretti, F., Gambini, A., & Santi, G. (2020). The Gestinv Database: A Tool for Enhancing Teachers Professional Development within a Community of Inquiry. In H. Borko and D. Potari (Eds.), *Proceedings of the Twenty-fifth ICMI Study School Teachers of mathematics working and learning in collaborative groups* (pp.621–628). Portugal: University of Lisbon.
- Ferretti, F., & Giberti, C. (2021). The properties of powers: Didactic contract and Gender Gap. *International Journal of Science and Mathematics Education*, 19(8), 1717-1735.
- Ferretti, F., Giberti, C., & Lemmo, A. (2018). The Didactic Contract to interpret some statistical evidence in mathematics standardized assessment tests. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(7), 2895-2906. DOI: <https://doi.org/10.29333/ejmste/90988>
- Ferretti, F., Lemmo, A. & Martignone, F. (2018a). La probabilità nelle prove INVALSI: analisi in verticale. *Induzioni*, 55, 27–47
- Ferretti, F., Lemmo, A., & Martignone, F. (2018b). Attained curriculum and external assessment in Italy: how to reflect on them. *Proceedings of the Twenty-fourth ICMI Study School Mathematics Curriculum Reforms: Challenges, Changes and Opportunities*, 381-388.
- Ferretti, F., Martignone, F., & Santi, G. . R. P. (2022). Analysis of standardized tests and pre-service teacher education: reflections on developed teachers' specialized knowledge. In Hodgen, J., Geraniou, E., Bolondi, G. & Ferretti, F. (Eds.), *Proceedings of the Twelfth Congress of European Research in Mathematics Education (CERME12)*. Bozen-Bolzano, Italy: ERME/Free University of Bozen-Bolzano.
- Ferretti, F., Michael-Chrysanthou, P., & Vannini, I. (2018). *Formative assessment for mathematics teaching and learning: Teacher professional development research by videoanalysis methodologies*. FrancoAngeli.
- Ferretti, F., Santi, G. R. P., & Bolondi, G. (2022). Interpreting difficulties in the learning of algebraic inequalities, as an emerging macro-phenomenon in Large Scale Assessment. *Research in Mathematics Education*, 1-23.
- Ferretti, F., Santi, G. R. P., Del Zozzo, A., Garzetti, M., & Bolondi, G. (2021). Assessment practices and beliefs: Teachers’ perspectives on assessment during long distance learning. *Education Sciences*, 11(6), 264.
- Gambini, A., De Simoni, M., & Ferretti, F. (2022). Predictive tools for university performance: an explorative study. *International Journal of Mathematical Education in Science and Technology*, 1-27.
- Giberti, C. (2018). Differenze di genere e misconcezioni nell’operare con le percentuali: Evidenze dalle prove INVALSI [Gender differences and misconceptions in working with percentages: Evidence from INVALSI tests]. *CADMO*, 2018(2), 97–114. <https://doi.org/10.3280/CAD2018-002007>
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational evaluation and policy analysis*, 11(3), 255-274.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and teaching*, 8(3), 381-391.

- Goldin, G. A., Hannula, M. S., Heyd-Metzuyanim, E., Raimo Kaasila, A. J., Lutovac, S., Di Martino, P., Morselli, F., Middleton, J. A., Pantziara, M., Zhang, Q. (2016). *Attitudes, beliefs, motivation and identity in mathematics education: An overview of the field and future directions*. Springer Nature.
- Grossman, P. L., Smagorinsky, P., & Valencia, S. (1999). Appropriating tools for teaching English: A theoretical framework for research on learning to teach. *American journal of Education*, 108(1), 1-29.
- Impedovo, M., Orlandoni, A., & Paola, D. (2011). *Quaderni SNV N. 1-MAT. Guida sintetica alla lettura della prova di Matematica*. Frascati: Invalsi.
- INVALSI (2017). Rilevazioni nazionali degli apprendimenti 2016-2017. Rapporto tecnico. Retrieved from https://www.invalsi.it/invalsi/doc_eventi/2017/Rapporto_tecnico_SNV_2017.pdf
- INVALSI (2018). Quadro di riferimento delle prove di INVALSI di matematica. Retrieved from https://invalsi-areaprove.cineca.it/docs/file/QdR_MATEMATICA.pdf
- Jaworski, B. (2006). Theory and practice in mathematics teaching development: Critical inquiry as a mode of learning in teaching. *Journal of mathematics teacher education*, 9(2), 187-211.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), 14-26.
- Kanes, C., Morgan, C. & Tsatsaroni, A. (2014). The PISA mathematics regime: knowledge structures and practices of the self. *Educational Studies in Mathematics* 87, 145–165.
- Kloosterman, P., Walcott, C., Brown, N. J. S., Mohr, D., Pérez, A., Dai, S., Roach, M., Dager Hall, L., & Huang, H. C. (2015). Using NAEP to Analyze Eighth-Grade Students' Ability to Reason Algebraically. In: Middleton, J., Cai, J., Hwang, S. (eds) *Large-Scale Studies in Mathematics Education*. Research in Mathematics Education. Springer, Cham. https://doi.org/10.1007/978-3-319-07716-1_9
- Lemmo, A., Branchetti, L., Ferretti, F., Maffia, A., & Martignone, F. (2015). Students' difficulties dealing with number line: a qualitative analysis of a question from national standardized assessment", *Quaderni di Ricerca in Didattica (Mathematics)*. In *CIEAEM: 20-24 Luglio* (Vol. 25, No. Supplement 2, pp. 143-150). GRIM.
- Looney J.W. (2011). Integrating Formative and Summative Assessment: Progress Toward a Seamless System?, *OECD Education Working Papers*, 58, OECD Publishing.
- Martignone, F. (2016). Un'attività di formazione per insegnanti di scuola secondaria di primo grado: analisi di prove Invalsi di matematica. *Form@ re*, 16(1).
- Martignone, F. (2017). Analysis of mathematics standardized tests: examples of tasks for teachers. In Dooley, T., & Gueudet, G. (Eds.), *Proceedings of the Tenth Congress of the European Society for Research in Mathematics Education* (pp. 3344–3351). Dublin, Ireland: DCU Institute of Education and ERME.
- Meinck, S., Neuschmidt, O., Taneva, M. (2017). Workshop Theme: "Use of Educational Large-Scale Assessment Data for Research on Mathematics Didactics". In: Kaiser, G. (eds) *Proceedings of the 13th International Congress on Mathematical Education*. ICME-13 Monographs. Springer, Cham. https://doi.org/10.1007/978-3-319-62597-3_132
- Morris, A. (2011). *Student Standardised Testing: Current Practices in OECD Countries and a Literature Review*. OECD Education Working Papers, No. 65, OECD Publishing.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research). Expanded edition (1980), with foreword and afterword by B. D. Wright. Chicago: University of Chicago Press.
- Radford, L. (2006). Elements of a cultural theory of objectification. In L. Radford, & B. D'Amore (Eds.), *Semiotics, culture and mathematical thinking [special issue]*. *Revista Latinoamericana de Investigación en Matemática Educativa (RELIME)* (pp. 103–129).
- Radford, L. (2021). *The theory of objectification: A Vygotskian perspective on knowing and becoming in mathematics teaching and learning*. Brill.

- Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coins. *Educational Studies in Mathematics*, 22, 1–36.
- Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 69(2), 107-131.
- Spagnolo, C., Capone, R., Ferretti, F., & Gambini, A. (2021). Quali strategie adottano gli studenti per leggere il testo di un problema? Primi risultati di uno studio di eye-tracking con quesiti INVALSI e OCSE-PISA. In P. Falzetti (a cura di), *I dati INVALSI come strumento per migliorare e valutare le competenze trasversali. IV Seminario "I dati INVALSI: uno strumento per la ricerca"* (pp. 68-99). Roma: Franco Angeli.
- Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*. Springer Nature. <https://doi.org/10.1007/978-3-319-32394-7>
- Tasaki, N. (2017). The impact of OECD-PISA results on Japanese educational policy. *European Journal of Education*, 52(2), 145-153.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage.